Formalisation et résolution de problèmes de Markov temporels par couplage avec VLE

Emmanuel Rachelson

Frédérick Garcia [2]
Patrick Fabiani [1]

[1]ONERA-DCSD

[2]INRA-BIA

Séminaire CD, 5 février 2008

Plan

Problèmes de Markov temporels

- Exemples
- Formalisation - le lien avec les MDP
- Une représentation plus compacte et structurée

Méthode de résolution . . .

- Difficultés liées aux problèmes de Markov temporels
- Principe de la méthode proposée

. . . et mise en pratique par couplage avec VLE

- Simuler des GSMP via VLE,
- ajouter une composante décisionnelle,
- et introduire une boucle simulation-optimisation

Plan

Problèmes de Markov temporels
- Exemples
- Formalisation - le lien avec les MDP
- Une représentation plus compacte et structurée

Méthode de résolution . . .
- Difficultés liées aux problèmes de Markov temporels
- Principe de la méthode proposée

. . . et mise en pratique par couplage avec VLE
- Simuler des GSMP via VLE,
- ajouter une composante décisionnelle,
- et introduire une boucle simulation-optimisation

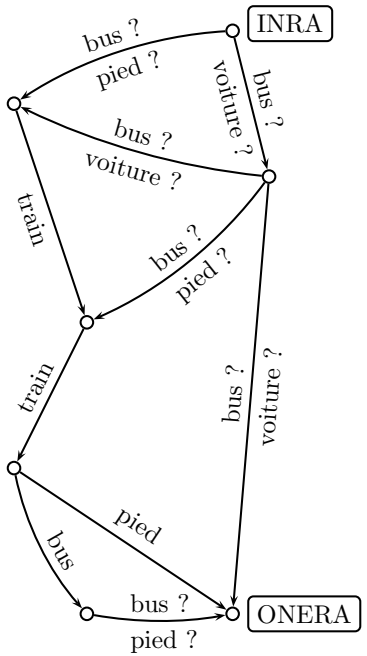
Exemples

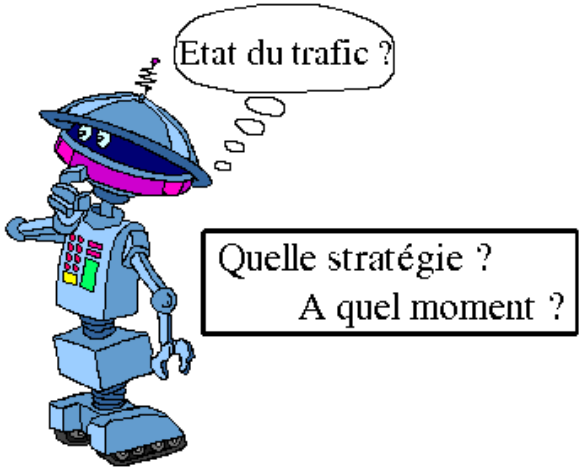
Décision dans l'incertain avec forte dynamique temporelle.
$\longrightarrow$ planifier pour se coordonner avec un environnement instationnaire et incertain.

Exemples

Décision dans l'incertain avec forte dynamique temporelle.
→ planifier pour se coordonner avec un environnement instationnaire et incertain.

Aller de l'INRA à l'ONERA ?





Exemples

Décision dans l'incertain avec forte dynamique temporelle.
$\longrightarrow$ planifier pour se coordonner avec un environnement instationnaire et incertain.

Quand ouvrir quel guichet ?

Exemples

Décision dans l'incertain avec forte dynamique temporelle.
$\longrightarrow$ planifier pour se coordonner avec un environnement instationnaire et incertain.

Routage des avions sur taxiway

Exemples

Décision dans l'incertain avec forte dynamique temporelle.
$\longrightarrow$ planifier pour se coordonner avec un environnement instationnaire et incertain.

Coordination à bord

Exemples

Décision dans l'incertain avec forte dynamique temporelle.
$\longrightarrow$ planifier pour se coordonner avec un environnement instationnaire et incertain.

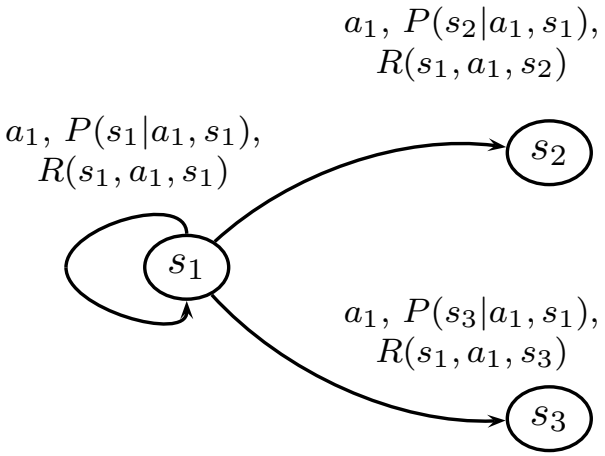
Ajouter ou retirer des rames ?

Processus stochastique

Problème de décision séquentielle

- Transitions stochastiques entre états décrits par des variables discrètes ou continues
- Composante temporelle

Cadre (X)MDP : $< S, A(X), p(s', t'|s, t, a(x)), r(s, t, a(x)) >$



XMDP : MDP + temps continu observable.

Caractéristiques des exemples

Pourquoi écrire un XMDP pour les exemples précédents est-il une tâche difficile ?

$\longrightarrow$ “il se passe beaucoup de choses complexes en parallèle”

- dynamique partiellement incontrolable
- phénomènes concurrents

Concurrence et (S)MDP

Semi-MP = SMP

Processus markoviens avec temps de séjour variable.

SMDP

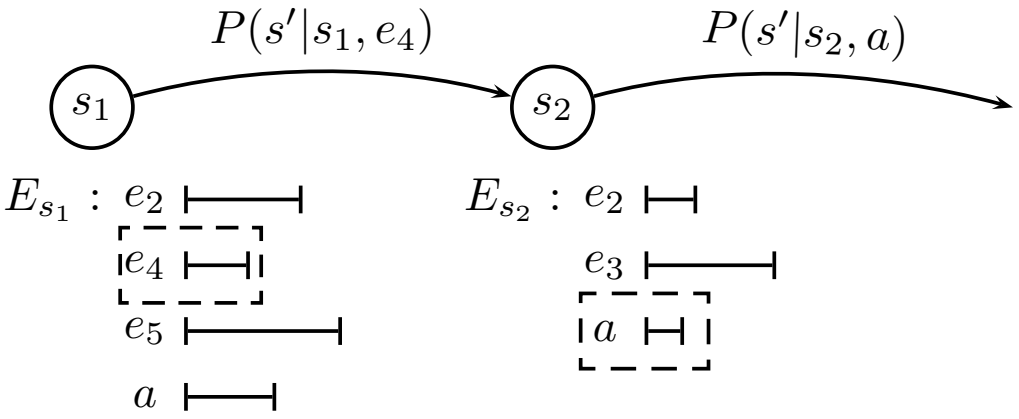
SMP + choix d'actions. Similaire aux MDP.

Generalized SMP = GSMP

SMP concurrents sur état commun. Evénements concurrents.

GSMDP

GSMP + choix d'actions.



Propriétés des GSMDP

Se ramènent à des MDP si les horloges sont observables ...
pb : les horloges ne sont pas observables.
$\longrightarrow$ dynamique non-Markovienne.

(Younes, 04) : approximation basée sur les distributions phase-type.
Introduit des états supplémentaires.

Proposition : résolution par insertion de t dans espace d'état, on cherche $\pi(s,t)$.

Plan

Problèmes de Markov temporels
- Exemples
- Formalisation - le lien avec les MDP
- Une représentation plus compacte et structurée

Méthode de résolution . . .
- Difficultés liées aux problèmes de Markov temporels
- Principe de la méthode proposée

. . . et mise en pratique par couplage avec VLE
- Simuler des GSMP via VLE,
- ajouter une composante décisionnelle,
- et introduire une boucle simulation-optimisation

Ce qui ne marche pas en pratique

- problème des variables continues :
  $\longrightarrow$ la résolution dépend fortement des hypothèses de représentation du modèle.
- problème de la taille de l'espace des trajectoires :
  $\longrightarrow$ curse of dimensionality sur variables discrètes,
  $\longrightarrow$ difficulté d'intégration des équations.
- problème d'une solution sous forme de politique :
  $\longrightarrow$ forme de la politique sur variables continues
  $\longrightarrow$ optimisation en ligne : mauvais comportement anytime de l'itération de la valeur
  $\longrightarrow$ optimisation en ligne : taille de la politique complète et pertinence de l'exploration

Ce qui ne marche pas en pratique

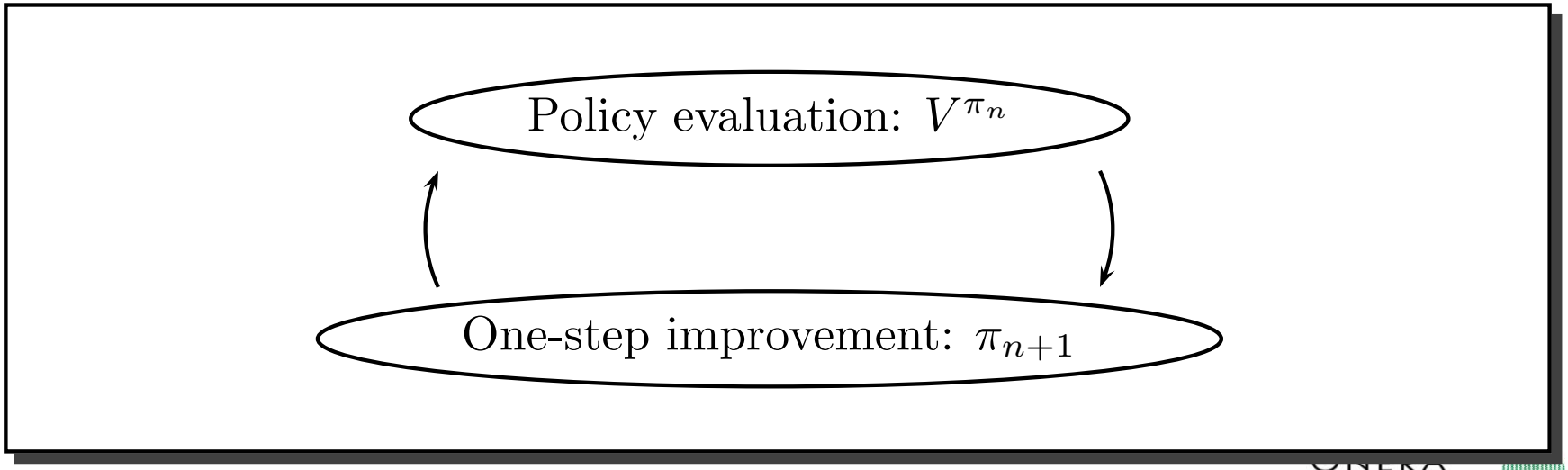
- problème des variables continues :
  $\longrightarrow$ la résolution dépend fortement des hypothèses de représentation du modèle.
- problème de la taille de l'espace des trajectoires :
  $\longrightarrow$ curse of dimensionality sur variables discrètes,
  $\longrightarrow$ difficulté d'intégration des équations.
- problème d'une solution sous forme de politique :
  $\longrightarrow$ forme de la politique sur variables continues
  $\longrightarrow$ optimisation en ligne : mauvais comportement anytime de l'itération de la valeur
  $\longrightarrow$ optimisation en ligne : taille de la politique complète et pertinence de l'exploration

Ce qui ne marche pas en pratique

- problème des variables continues :
  $\longrightarrow$ la résolution dépend fortement des hypothèses de représentation du modèle.
- problème de la taille de l'espace des trajectoires :
  $\longrightarrow$ curse of dimensionality sur variables discrètes,
  $\longrightarrow$ difficulté d'intégration des équations.
- problème d'une solution sous forme de politique :
  $\longrightarrow$ forme de la politique sur variables continues
  $\longrightarrow$ optimisation en ligne : mauvais comportement anytime de l'itération de la valeur
  $\longrightarrow$ optimisation en ligne : taille de la politique complète et pertinence de l'exploration

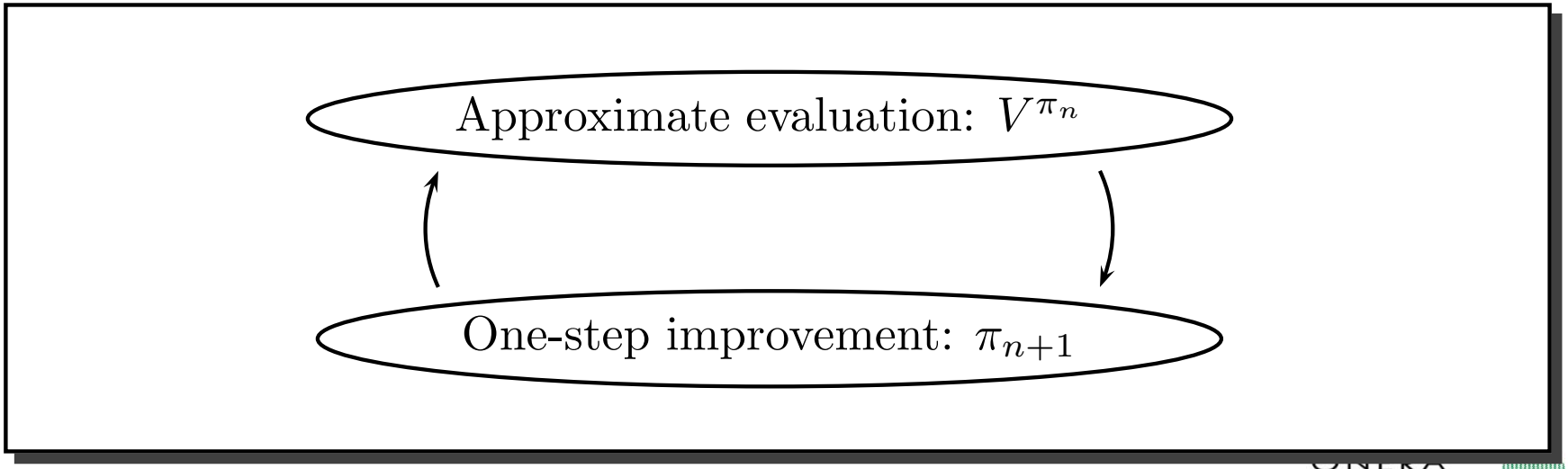
online-ATPI

- Itération de la politique approchée
- Phase d'évaluation ? Simulation de la politique
- Généralisation de la fonction de valeur par apprentissage à partir des simulations
- Stockage de la politique ? Instanciation en ligne



online-ATPI

- Itération de la politique approchée
- Phase d'évaluation ? Simulation de la politique
- Généralisation de la fonction de valeur par apprentissage à partir des simulations
- Stockage de la politique ? Instanciation en ligne



online-ATPI

- Itération de la politique approchée
- Phase d'évaluation ? Simulation de la politique
- Généralisation de la fonction de valeur par apprentissage à partir des simulations
- Stockage de la politique ? Instanciation en ligne

Intérêt : recherche monotone dans l'espace des politiques,
on évite les minima locaux.
$\longrightarrow$ bon comportement anytime.

online-ATPI

- Itération de la politique approchée
- Phase d'évaluation ? Simulation de la politique
- Généralisation de la fonction de valeur par apprentissage à partir des simulations
- Stockage de la politique ? Instanciation en ligne

Echantillonage dans l'espace des trajectoires
$\longrightarrow$ simulation basée sur VLE permet d'envisager le couplage de modèles hétérogènes.

online-ATPI

- Itération de la politique approchée
- Phase d'évaluation ? Simulation de la politique
- Généralisation de la fonction de valeur par apprentissage à partir des simulations
- Stockage de la politique ? Instanciation en ligne

$\longrightarrow$ Méthode utilisée : SVM.

A envisager : réseaux de neurones, fonctions de voisinage.

online-ATPI

- Itération de la politique approchée
- Phase d'évaluation ? Simulation de la politique
- Généralisation de la fonction de valeur par apprentissage à partir des simulations
- Stockage de la politique ? Instanciation en ligne

$\longrightarrow$ Calcul de l'action optimale par roll-out et échantillonage sur un coup en utilisant la fonction de valeur apprise

Plan

Problèmes de Markov temporels
- Exemples
- Formalisation - le lien avec les MDP
- Une représentation plus compacte et structurée

Méthode de résolution . . .
- Difficultés liées aux problèmes de Markov temporels
- Principe de la méthode proposée

. . . et mise en pratique par couplage avec VLE
- Simuler des GSMP via VLE,
- ajouter une composante décisionnelle,
- et introduire une boucle simulation-optimisation

L'extension GSMP

Objectif : simuler des GSMP dans VLE

Interface : l'utilisateur fournit
- une liste de variables d'état,
- une liste d'événements,

puis spécifie $P(s'|s,e,c_e)$, $F(c|s,e)$ et E_s.

L'extension GSMP

```
double GSMP_metro::arr_pass_0_duration::operator()
            (const GSMP::state_map & /* state */,
             const double & time)
{
    double average = 3*cos(2 * 3.1415 / (60*12.) * time) + 4;
    double std_dev = 0.8*cos(2 * 3.1415 / (60*12.) * time) + 0;
    double r = -1;
    while(r<=0){r=utils::Rand::rand().normal(std_dev, average);}
    return r;
}

void GSMP_metro::arr_pass_0_effect::operator()
            (GSMP::state_map & state,
             const std::pair<double,double> & /* time_pair */)
{
    std::string quai = "quai_0";
    if(state.get_discrete_variable(quai)<50)
    {
        state.set_discrete_variable(quai,
                                    state.get_discrete_variable(quai)+1);
    }
}
```

L'extension GSMP

En interne, à chaque transition, δ_{int} . . .

- Active et désactive les événements du GSMP,
- Met à jour les horloges des événements,
- Tire l'état suivant.

L'extension GSMDP

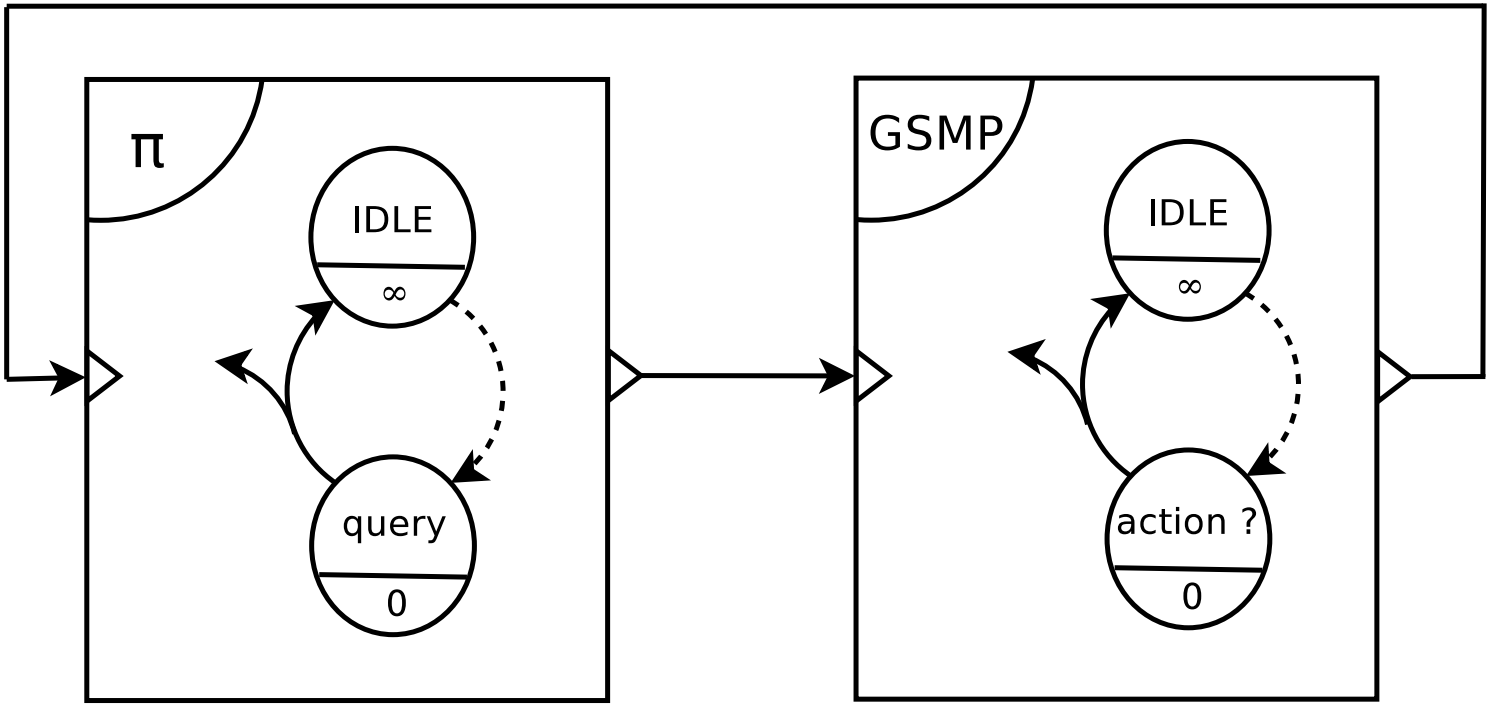
Objectif : Simuler un GSMDP contrôlé par une politique spécifiée par l'utilisateur

Interface : L'utilisateur fournit
- une liste de variables d'état,
- une liste d'événements,
- une liste d'actions,

puis spécifie $P(s'|s,e,c_e)$, $F(c|s,e)$, E_s et $\pi(s)$.

L'extension GSMDP

Schéma simplifié des modèles DEVS



Algorithme général

```
init : π0
répéter :
    base_app = ∅
    répéter n fois
        trajectoire = simu(V (ou π0 à défaut))
        base_app = base_app ∪ trajectoire
    V = SVM(base_app)
```

Conclusion

Développement dans VLE

- Deux extensions finies, à améliorer.
- Utilisation du plugin Storage pour la mise en place de la boucle optimisation-simulation.

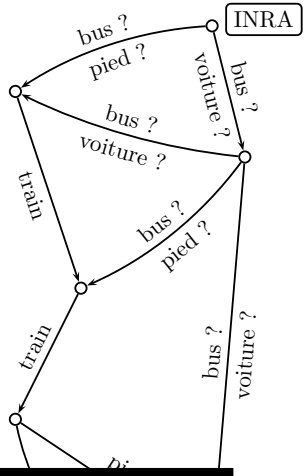
Amélioration de l'algorithme

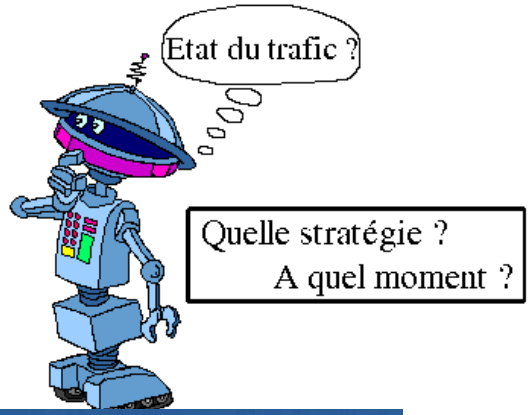
- Autres techniques de régression
- Raffinement incrémental de V en modifiant les conditions initiales de simulation pendant la phase d'évaluation de π

Bilan

- Communauté planification : base de modèles VLE compatibles ?
- Exemple d'une implantation du couplage VLE-planif.

ONERA THE FRENCH AEROSPACE LAB INRA





Questions ?

ONERA THE FRENCH AEROSPACE LAB INRA